No. 150 (82-17)

Exact rate $n^{-2/3}$ in the empirical Bayes estimation: case of uniform distributions I.

by

Yoshiko Nogami

April 1982

1. Introduction

The empirical Bayes problem is first introduced by Robbins (1955). Fox (1970, 1978) and Susarla and O'Bryan (1979) so far considered the empirical Bayes estimation problem for the case of uniform distributions. Besides them, Nogami (1978, 1979, 1981, 1982) considered the squared error loss estimation (SELE) in the set-compound problem under the family of certain retracted distributions including uniform distributions as a special case. This paper is a continuation of Fox's work (1978) and we consider the SELE of θ under the uniform distribution $U[\theta, \theta+1)$ on the interval $[\theta, \theta+1)$ for $\theta \in \Omega = [c, d]$ where $-\infty < c < d < \infty$.

As notational conventions we use the following devices. A distribution function H will also be used to denote the associated measure. The argument of a function will not be displayed sometimes and operator notation will be used to represent integrals, e.g., $\int g(t)\, d\mu(t)$ might be expressed as $\mu(g(t))$ or $\mu(g)$. [A] denotes the indicator function of the event A. y-1 often abbreviates y'. $\doteq$ means a defining property.

Let $f_\theta(x) = [\theta \le x < \theta + 1]$. Let G be a (unknown) prior distribution on Ω. Defining the c.d.f. of above f_θ by F_θ we let $F(x) = \int F_\theta(x)\, dG(\theta)$ and $f(x) = \int f_\theta(x)\, dG(\theta)$, i.e. F and f are respectively the marginal c.d.f. and p.d.f. of a random variable X. Let $(X_1, X_2, \ldots, X_n)$ be n i.i.d. past observations distributed according to F. Let X denote the (n+1)st observation X_{n+1}. Let P be the product measure on the space of $(X_1, \ldots, X_n, (\theta, X))$, resulting from F^n and the joint distribution of (θ, X). Let $\phi_G(x)$ denote the Bayes estimator vs G given by

$$(1.1)\qquad \phi_G(x) = \int \theta f_\theta(x)\, dG(\theta) / \int f_\theta(x) dG(\theta) = \int_{x'+}^{x} \theta\, dG/(G(x)-G(x'))$$

where 0/0 is defined to be x and the affix + is intended to describe the integration as over (x', x].

The risk of an empirical Bayes (EB) estimator t_n for θ is $R(t_n, G) = P((t_n(x) - \theta)^2)$. The Bayes estimator vs G achieves the Bayes envelope $R \doteq R(\phi_G, G) = \inf_\phi R(\phi, G)$. We call t_n asymptotically optimal (a. O.) when $R_n - R \to 0$ as $n \to \infty$. Since when $R(t_n, G)$ and R are finite we have

$$(1.2) \qquad 0 \leq R(t_n, G) - R = P(t_n(X) - \phi_G(X))^2,$$

our purpose is to find rates of convergence of $P(t_n - \phi_G)^2$ to zero.

Let ϕ_n be the EB estimator proposed by Fox (1978, Section 3) (where there is a misprint and $\phi_n(x) = \phi(x)$ should be of form $x' + \psi(x)[k(x) > 0]$) and ϕ_n^* be the EB estimator presented by Nogami (1981), adopted in the EB problem. In this paper we shall show that under certain assumptions for G (see below Ai) $\sim$ Aiii)) ϕ_n and ϕ_n^* are both a.O. with the exact order $n^{-2/3}$ of convergence. However, from the structure of construction ϕ_n cannot be extended to a more general family of certain retracted distributions to the interval $[\theta, \theta+1)$ as, for example, in Nogami (1981), but ϕ_n^* can do so. But, for the sake of simplicity we shall present bounds for $R(\phi_n, G) - R$ in Sections 3 and 4 and by a usage of the methods in Section 3 and a lower bound in Nogami (1981) we obtain the exact rate for $R(\phi_n^*, G) - R$ in Section 5.

Let E_x and E be the conditional product measure on the space of $(X_1, \ldots, X_n, \theta | x)$ and the marginal probability measure of X.

Assumptions for G:

Ai) The support of G is bounded, i.e.

$\Omega = [c, d]$ where $-\infty < c < d < +\infty$.

Aii) For some $(1 >) h > 0$, $\sup_{s\in[c,d+1]} \ \sup_{\substack{\text{a.e.} \\ y\in(s,s+h)}} G'(y) < M(< \infty)$.

Aiii) $(0 <) \int_c^{d+1} (f(y))^{-1} \, dy \leq K \ (< \infty)$.

We denote $R(\phi_n, G)$ and $R(\phi_n^*, G)$ by R_n and R_n^*, respectively. Let $\vee$ and $\wedge$ denote the supremum and the infimum, respectively.

2. Construction of ϕ_n.

Fix $X = x$ where x is a realization of X. Since $F(x) = G(x') + xf(x) - \int_{x'+}^{x} \theta \, dG(\theta)$, ϕ_G in (1.1) is written by

(2.1) $$\phi_G(x) = x - \psi(x)$$

where

(2.2) $$\psi(x) = (F(x) - G(x'))/f(x).$$

Since the conditional distribution of θ given x is concentrated on $(x', x]$, $0 \leq \psi \leq 1$. Note that $P([f(X) > 0]) = 1$. Let $F_n(y) = n^{-1} \sum_{j=1}^{n} [X_j \leq y]$ and $\hat{f}(y) = h^{-1}(F_n(y+h) - F_n(y))$ where h is chosen so that $0 < h < 1$. Since $f(y) = G(y) - G(y')$,

(2.3) $$G(y) = \sum_{r=0}^{\infty} f(y-r)$$

where the number of r to sum is finite. Hence, we estimate G(y) by

$$(2.4)\qquad G^*(y) = \sum_{r=0}^{\infty} \hat{f}(y-r).$$

We also estimate F(y) by the empiric distribution $F_n(y)$ of n observations $X_1, X_2, \ldots, X_n$. Thus, in view of (2.1) and (2.2), Fox's estimate $\phi_n(x)$ (at X=x) for θ is given by

$$(2.5)\qquad \phi_n(x) = x - (0 \vee \psi_n(x)) \wedge 1$$

where

$$(2.6)\qquad \psi_n(x) = (F_n(x) - G^*(x'))/\hat{f}(x).$$

In the next section we shall find an upper bound for $R_n - R$.

3. An Upper bound for $R_n - R$.

In this section we use Lemma 4.1 of Singh (1979) and obtain a rate $0(n^{-2/3})$ for $R_n - R$ with a choice of $h = n^{-1/3}$.

In (2.2) and (2.6), let $\Psi(X) = u/w$ and $\psi_n(X) = U/W$. In view of (1.2) with t_n replaced by ϕ_n,

$$(3.1)\qquad 0 \le R_n - R = E\{E_X(|\frac{U}{W} - \frac{u}{w}| \wedge 1|^2)\}.$$

For X = x given, Lemma 4.1 of Singh (1979) leads to the inequality

$$(3.2)\qquad E_X(|\frac{U}{W} - \frac{u}{w}|\wedge)^2 \le 1 \wedge A \qquad \text{a.e.E}$$

where

(3.8) $$(nh^2)\ \mathrm{Var}_X(\hat{f}(v)) \le n^{-1} \sum_{j=1}^{n} E_X Y_j \le h$$

where the last inequality follows because $f(\cdot) \le 1$.

To bound the second term of rhs(3.7) we have by transformation theorem and the triangular inequality that

(3.9) $$|E_X \hat{f}(v) - f(v)| = |\int_0^1 (f(v+hy) - f(v))\, dy|$$

$$\le \int_0^1 \{G(v+hy) - G(v)\}dy + \int_0^1 \{G(v'+hy) - G(v')\}dy.$$

Since G is of bounded variation on $[v, v+\eta]$ (cf. Royden (1968)) with $\eta = hy$, there exists the first derivative G' a.e. in $(v, v+\eta)$. Furthermore, $G(z)$ is right continuous. These facts together with our assumption Aii) satisfy the requirements for the following Taylor expansion (cf. Singh (1978, p.639)):

$$G(v+hy) = G(v) + \int_v^{v+hy} G'(z)\, dz.$$

Above equality still holds even if we replace v by v'. Thus, by applying these to the extreme rhs(3.9) and using Aii),

$$|E_v\ \hat{f}(v) - f(v)| \le \int_0^1 \int_v^{v+hy} G'(z)\, dzdy + \int_0^1 \int_{v'}^{v'+hy} G'(z)\, dzdy \le Mh.$$

This, (3.8) and (3.7) give us the bound of the lemma.

Let $c_0, c_1, c_2, \ldots, c_7$ be positive constants. By (3.4), (3.5), Lemma 3.1 and weakening the resulted bound, $E_X|u-U|^2 \le c_0(nh)^{-1} + c_1h^2$. Similarly, $E_X|w-W|^2 \le c_2(nh)^{-1} + c_3h^2$. Thus, lhs(3.2) $\le (c_4(nh)^{-1} + c_5h^2)|w|^{-2}$ because $0 \le u/w \le 1$. Since $E(|w|^{-2}) \le K(< +\infty)$ from our

(3.3) $$A = 2^3|w|^{-2}\{E_x|u-U|^2 + ((u/w)^2 + 2^{-1})E_x|w-W|^2\}.$$

Let N be the greatest integer less than d+2-c. In view of (2.3) and the definition of G*, applying the c_r-inequality (Loève (1963, p.152)) (N+1) times leads to

(3.4) $$2^{-1}E_X|u-U|^2 \leq E_X|F(x)-F_n(x)|^2 + \sum_{r=1}^{N} 2^r E_X|f(x-r) - \hat{f}(x-r)|^2\}.$$

Since $E_X(F_n(x)) = F(x)$, the first term of rhs(3.4) is $Var_X(F_n(x))$, the variance of $F_n(x)$. Because $F_n(x)$ is the average of n i.i.d. random variables with the same variance $F(x)(1-F(x))$,

(3.5) $$(\text{the first term of rhs}(3.4)) = n^{-1}(F(x)(1-F(x))) \leq n^1.$$

To get bounds for the second term of rhs(3.4) and $E_x|w-W|^2$ we use Lemma 3.1 below.

Lemma 3.1 For each $r \in \{1, 2, \ldots, N\}$, fixed, let $v = x-r$. Then, with $M(>0)$ in Aii)

(3.6) $$2^{-1} E_x|f(v) - \hat{f}(v)|^2 \leq (nh)^{-1} + M^2h^2.$$

Proof By c_r-inequality (Loève (1962)),

(3.7) $$\text{lhs}(3.6) \leq Var_X(\hat{f}(v)) + (E_X(\hat{f}(v)) - f(v))^2.$$

To get a bound for the first term of rhs(3.7) we let $Y_j = [v < x_j \leq v+h]$, j=1, 2, ..., n. Since $\hat{f}(v) = (nh)^{-1} \Sigma_{j=1}^n Y_j$, Y_j's are independent and $E_XY_j = \int_v^{v+h} f(z)dz$, assumption Aiii), in view of (3.1) we obtain the following theorem:

Theorem 3.1 With a unknown prior G satisfying assumptions Ai), Aii) and Aiii),

(3.10) $$0 \le R_n - R \le c_6(nh)^{-1} + c_7h^2.$$

From above Theorem 3.1 we can see that with a choice of $h = n^{1/3}$, $R_n - R = O(n^{-2/3})$.

4. A lower bound for R_n-R and the exact rate $n^{-2/3}$ for ϕ_n.

In this section we assume that G is a degenerate distribution function at some point $\theta_0 \in \Omega$. Furthermore, without loss of generality we assume $\theta_0 \equiv 0$. Then, $\phi_G(x) \equiv 0$ and thus

(4.1) $$R_n - R = P(\phi_n^2).$$

In Section 3 we have shown that with $h = n^{-1/3}$, $R_n - R = O(n^{-2/3})$. Let k_0 and k_1 be positive constants. In this section we shall obtain Theorem 4.1 where for $h=h_n$ where $nh \to \infty$ and $h \to 0$, $R_n - R \ge k_0(nh)^{-1}$. Hence, combining these results together we shall see that for sufficiently large n and a choice of $h = n^{-1/3}$,

(4.2) $$k_0 n^{-2/3} \le R_n - R \le k_1 n^{-2/3}.$$

For $X_{n+1} = x$ fixed, let $\hat{\psi}_n(x) = x - \psi_n(x)$ in (2.6). Letting $u = \sum_{j=1}^{n} [0 \le X_j \le x]$ and $v = \sum_{j=1}^{n} [x < X_j \le x+h]$ we have

$$(4.3)\qquad \hat{\psi}_n(x) = v^{-1}\{xv - hu + \sum_{j=1}^{n} [0 < X_j \leq x'+h]\} \qquad \text{a.e.}E_x.$$

Note that in (2.5) $\phi_n = x' \vee \psi_n$ for $x \in [0, 1-h)$; $= (x' \vee \psi_n) \wedge x$ for $x \in [1-h, 1)$.

Let $B = [\psi_n \geq x', x < 1-h]$. By (4.3) and the difinition (2.5) of ϕ_n

$$(4.4)\qquad P\phi_n^2 \geq P(\hat{\psi}_n^2\, B).$$

Define $Y = (nx(1-x))^{-1/2}(u-nx)$ and $Z = (nh)^{-1/2}(v-nh)$. Then,

$$(4.5)\qquad \hat{\psi}_n(x) = \frac{(nh)^{-1/2}xZ}{1+(nh)^{-1/2}Z} - \frac{\sqrt{x(1-x)}\; n^{-1/2}Y}{1+(nh)^{-1/2}Z}.$$

To prove Theorem 4.1 below we use Lemmas 2.2 and 2.4 of Nogami (1981) which are furnished to get a lower bound of the modified regret for the estimate ϕ^* for $\theta_1 = \ldots = \theta_n = 0$. For convenience we write the explicit form of ϕ^* as follows:

$$(4.6)\qquad \phi^*(x) = (x' \vee \psi(x)) \wedge x$$

where for every $x \in [0, 1)$

$$(4.7)\qquad \psi(x) = \{\sum_{j=1}^{n} (X_j - h)[x < X_j \leq x+h] - h\sum_{j=1}^{n}[0 < X_j \leq x] - h + \sum_{j=1}^{n}[0 \leq X_j \leq x'+h]\} / \sum_{j=1}^{n}[x < X_j \leq x+h] \qquad \text{a.e.}E_x$$

<u>Theorem 4.1</u>. If h is a function of n such that $nh \to \infty$ and $h \to 0$, then for any $0 < \varepsilon < \frac{1}{3}$, there exists $N_0 < +\infty$ so that for all $n \geq N_0$,

$$R_n - R > \left(\frac{1}{3} - \varepsilon\right)\frac{1}{nh}.$$

Proof Fix $x \in (0.1)$ until (4.8). In view of (4.3) and (4.7), $\psi \leq \hat{\psi}_n$. Hence by Lemma 2.4 of Nogami (1981),

(4.8) $$E_X[\hat{\psi}_n \leq x] \leq E_x[\psi \leq x] \to 0 \quad \text{for given } x$$

Let $\overset{\mathcal{D}}{\to}$ and $\overset{P}{\to}$ denote convergence in distribution and convergence in probability respectively. Also, $N(a, b)$ denotes the normal distribution with mean a and variance b. Since by Lemma 2.2 of Nogami (1981) $(Y, Z) \overset{\mathcal{D}}{\to} N(\underline{0}, I)$ where $\underline{0}$ is 2 dimensional zero vector and I, 2×2 identity matrix, and since by (4.8) $B \overset{P}{\to} 1$, it follows from Slutsky's theorem applied to rhs(4.5) that if $x \in (0, 1)$, then

$$\sqrt{nh}\, \hat{\psi}_n B \quad \overset{\mathcal{D}}{\to} \quad N(0, x^2).$$

As a consequence of a convergence theorem (cf. Loéve (1963, 11.4, A(i))) we have

(4.9) $$\underline{\lim}\ (nh)\ E_x(\hat{\psi}_n^2 B) \geq x^2[0 < x < 1].$$

Thus, by Fatou's Lemma applied to the lhs below

$$\underline{\lim}\ EE_x((nh)\hat{\psi}_n^2 B) \geq P(\text{lhs}(4.9)) \geq \int_0^1 y^2 dy = \frac{1}{3}.$$

Therefore, in view of (4.1) and (4.4),

$$\underline{\lim}\ (nh)(R_n - R) \geq 3^{-1}$$

and the definition of lim inf leads to the conclusion.

Theorems 3.1 and 4.1 leads to (4.2).

5. Exact rate $n^{-2/3}$ for $R_n^* - R$.

We introduce another EB estimate ϕ_n^* with the exact order $n^{-2/3}$ of convergence for $R_n^* - R$. The construction of the ϕ_n^* is similar to the set compound estimator ϕ^* presented by Nogami (1981). ϕ_n^* coincides with ϕ^* as in (4.6) when $X_1, \ldots, X_{n+1}$ are i.i.d. with $U[0, 1)$. Hence, lower bounds in Theorem 2.1 in Nogami (1981) apply for those for $R_n^* - R$.

Since $\phi_G(x)$ in (1.1) is alternatively written as

$$\phi_G(x) = x - \{(\int_0^1 G(x'+t)dt - G(x'))/f(x)\},$$

estimating G by G* in (2.4) and f by $\hat{f}$ we obtain another EB estimate $\phi_n^*(x)$ given by

$$\phi_n^*(x) = x - (0 \vee \psi_n^*(x)) \wedge 1$$

where

$$\psi_n^*(x) = \{\int_0^1 G^*(x'+t)dt - G^*(x')\}/\hat{f}(x).$$

In view of (2.5) we can see that the only difference between ϕ_n and ϕ_n^* is the first terms of the numerators of respective ψ_n and ψ_n^*.

To get an upper bound for $R_n^* - R$, we can proceed in the same way as we have done in Section 3 for $R_n - R$. However, the first term of rhs(3.4) is now replaced by

$$\int_0^1 \sum_{r=1}^{N} 2^r\{E_X(|\hat{f}(x+t-r) - f(X+t-r)|^2)\}dt$$

by a usage of Fubini theorem and N usages of c_r-inequality (Loéve (1963, p.152)). Thus, applying Lemma 3.1 we can easily obtain that

(5.1) $$0 \leq R_n^* - R_n \leq a_0(nh)^{-1} + a_1h^2$$

where a_0 and a_1 are positive constants.

A lower bound for $R_n^* - R$ is obtained from Theorem 2.1(ii) of Nogami (1981); If h is a function of n such that $nh \to \infty$, $h \to 0$ and $nh^3 = O(1)$, then for any $0 < \varepsilon < 3^{-1}$, there exists $N_0 < +\infty$ so that for all $n \geq N_0$

$$R_n^* - R > (3^{-1}-\varepsilon)(nh)^{-1}.$$

Therefore, by above inequality and (5.1) we obtain with $h = n^{-1/3}$ and for positive constants a_2 and a_3, $a_2\, n^{-2/3} \leq R_n^* - R \geq a_3\, n^{-2/3}$.

REFERENCES

[1] Fox, Richard J. (1970). Estimating the empiric distribution function of certain parameter sequences. Ann. Math. Statist. 41, 1845-1852.

[2] ———— (1978). Solutions to empirical Bayes squared error loss estimation problems. Ann. Statist. 6, 846-853.

[3] Loève, M. (1963). Probability Theory (3rd ed.), Van Nostrand, Princeton.

[4] Nogami, Y. (1978). The set compound one-stage estimation in the nonregular family of distributions over the interval (0, θ). Ann. Inst. Statist. Math., 30, A, 35-43.

[5] ———— (1979). The k-extended set-compound estimation problem in a nonregular family of distributions over [θ, θ+1). Ann. Inst. Statist. Math., 31, A, 169-176.

[6] ———— (1981). The set-compound one-stage estimation in the nonregular family of distributions over the interval [θ, θ+1). Ann. Inst. Statist. Math., 33, A, 67-80.

[7] ———— (1982). A rate of convergence for the set-compound estimation in a family of certain retracted distributions. Ann. Inst. Statist. Math., 34, A (To appear).

[8] Robbins, Herbert (1955). An empirical Bayes approach to statistics. Proc. Third Berkeley Symp. Math. Statist. Probability 1, 157-163.

[9] Royden, H. L. (1968). Real Analysis (2nd ed.). Macmillan Company, New York.

[10] Singh, R. S. (1978). Nonparametric estimation of derivatives of average of μ-densities with convergence rates and applications. SIAM J. Appl. Math., 35, No.4, 637-649.

[11] ———— (1979). Empirical Bayes estimation in Lebesgue-exponential families with rates near the best possible rate. Ann. Statist. 7, 890-902.

[12] Susarla, V. and O'Bryan, T. (1979). Empirical Bayes interval estimates involving uniform distributions. Commun. Statist. - Theor. Meth. A 8(4), 385-397.